

KLASIFIKASI CALON PENERIMA BIDIKMISI DENGAN MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

¹Hilda Nur Zerlinda, ²Isnandar Slamet, ³Etik Zukhronah

^{1,2,3} Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sebelas Maret
e-mail: zerlindaa10@gmail.com

Abstrak

Program beasiswa bidikmisi merupakan program bantuan yang diberikan oleh Kementerian Riset Teknologi dan Pendidikan Tinggi untuk peserta didik yang kurang mampu dalam ekonomi namun memiliki potensi yang baik dalam melanjutkan pendidikan ke jenjang yang lebih tinggi, yaitu perguruan tinggi. Pendaftar bidikmisi di UNS selalu mengalami peningkatan setiap tahunnya, namun kuota peserta yang diterima tidak sebanding dengan jumlah pendaftar. Penelitian ini bertujuan untuk mengklasifikasikan calon penerima bidikmisi UNS menggunakan algoritma *K-Nearest Neighbor*. Pada penelitian ini, digunakan sebanyak 2039 data pendaftar bidikmisi dengan beberapa variabel independen, seperti penghasilan orang tua, pekerjaan ayah dan ibu, jumlah tanggungan, luas tanah, sumber air, dan kepemilikan rumah. Tahap yang dilakukan pada penelitian ini yaitu menormalisasikan data, membagi data menjadi data *training* dan data *testing* serta menentukan nilai *k*. Berdasarkan hasil perhitungan dengan menggunakan algoritma *K-NN*, diperoleh nilai akurasi sebesar 84.4% dan nilai kappa sebesar 0.63 yang berarti bahwa hasil klasifikasi sudah sangat bagus.

Kata Kunci: Bidikmisi, Klasifikasi, KNN, Nilai akurasi

Abstract

The Bidikmisi scholarship program is an assistance program provided by the Ministry of Research, Technology and Higher Education for students who are less able in the economy but have good potential in continuing their education to a higher level, which is university. Bidikmisi registrants in the UNS always increase every year, but the quota of participants received is not proportional to the number of registrants. This study aims to classify UNS bidikmisi recipients using *K-Nearest Neighbor* algorithm. In this study, 2039 data on bidikmisi registrants were used with several independent variables, such as parents' income, father and mother's work, number of dependents, land area, water source, and home ownership. The steps taken in this research are normalizing data, dividing data into training data and testing data and determining the value of *k*. Based on the results of calculations using the *K-NN* algorithm, obtained an accuracy value of 84.4% and a kappa value of 0.63 which means that the classification result is very good.

Keywords: Bidikmisi, Classification, KNN, Accuracy value

PENDAHULUAN

Beasiswa adalah pemberian bantuan berupa keuangan yang berasal dari pemerintah bagi individu yang bertujuan untuk membantu keberlangsungan pendidikan yang ditempuh dan pemberiannya dilakukan secara selektif sesuai dengan syarat dari beasiswa yang ditawarkan. Terdapat beberapa jenis beasiswa yang biasanya ditawarkan oleh perguruan tinggi di Indonesia, seperti Beasiswa PPA (Peningkatan Prestasi Akademik), Beasiswa Toyota, Beasiswa Bank Indonesia, Beasiswa Bidikmisi, dan masih banyak lagi. Berbagai macam beasiswa tersebut memiliki kriteria berbeda satu sama lainnya yang harus dipenuhi oleh pendaftar. Salah satu beasiswa yang ditawarkan oleh Universitas Sebelas Maret adalah bidikmisi.

Bidikmisi merupakan program bantuan dari pemerintah yang ditujukan untuk mahasiswa kurang mampu namun mempunyai potensi akademik yang baik untuk melanjutkan pendidikan ke jenjang perguruan tinggi. Program ini memberikan fasilitas pembebasan biaya pendidikan serta subsidi biaya hidup. Beasiswa bidikmisi diselenggarakan setiap tahunnya untuk mahasiswa baru dengan kriteria penerima yang telah ditentukan oleh Kementerian Riset Teknologi dan Pendidikan Tinggi. Pada tahun ajaran 2018/2019, pendaftar bidikmisi di UNS tercatat sebanyak kurang lebih 2000 mahasiswa dengan kuota penerima bidikmisi kurang dari 1000 mahasiswa.

Oleh karena pendaftar yang selalu bertambah setiap tahunnya dan terbatasnya kuota mahasiswa yang diterima, maka diperlukan metode yang tepat untuk membantu proses seleksi mahasiswa pendaftar

bidikmisi. Juliardi (2017) melakukan klasifikasi mahasiswa penerima bidikmisi dengan menggunakan C4.5, didapatkan nilai akurasi sebesar 79.80%. Putri (2014) melakukan perbandingan algoritma *k-NN* dengan *Naive Bayes Classifier*, diperoleh hasil bahwa tingkat akurasi algoritma *k-NN* lebih besar yaitu 96.06%. Aulia dkk. (2015) melakukan perbandingan *k-NN* dengan *Support Vector Machine (SVM)*, didapatkan kesimpulan bahwa algoritma *k-NN* memiliki tingkat akurasi yang lebih besar yaitu sebesar 65%. Yunita (2017) melakukan perbandingan algoritma *k-NN* dengan *Decision Tree*, didapatkan kesimpulan bahwa *k-NN* memiliki tingkat akurasi yang lebih besar yaitu 98.18%.

Berdasarkan beberapa penelitian yang telah dipaparkan di atas, maka pada penelitian ini akan digunakan algoritma *k-NN* dalam mengklasifikasikan calon penerima bidikmisi. *K-NN* adalah suatu metode yang menggunakan algoritma *unsupervised* dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *k-NN* (Liu, 2007).

Bidikmisi

Program bidikmisi memiliki beberapa tujuan, diantaranya adalah (1) Meningkatkan akses dan kesempatan belajar di PT bagi peserta didik yang tidak mampu secara ekonomi dan memiliki prestasi akademik yang baik; (2) Meningkatkan prestasi mahasiswa, baik akademik maupun non akademik; dan (3) Melahirkan lulusan yang mampu berperan dalam upaya pemutusan mata rantai kemiskinan dan pemberdayaan masyarakat. Terdapat beberapa kriteria yang harus dipenuhi pendaftar agar bisa mendapatkan bidikmisi, salah satunya adalah total penghasilan orang tua tidak lebih dari Rp 3.000.000/bulan atau Rp 750.000 per anggota keluarga. Program ini memberikan fasilitas pembebasan biaya pendidikan dan subsidi hidup, yang diberikan kepada penerimanya selama delapan semester untuk S1/D4, enam semester untuk D3, empat semester untuk D2, dan dua semester untuk D1.

Algoritma *k-NN*

Algoritma *k-NN* adalah algoritma yang menentukan nilai jarak pada pengujian *data testing* dengan *data training* berdasarkan nilai terkecil dari nilai ketetanggaan terdekat. Tujuan dari algoritma *k-NN* adalah untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples*. Algoritma *k-NN* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Jarak yang digunakan adalah jarak *Euclidean* yang didefinisikan sebagai berikut,

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

dengan:

$d(x_i, x_j)$: jarak *Euclidean*
 (x_i) : *record* ke-*i*
 (x_j) : *record* ke-*j*
 a_r : data ke-*r*

Confusion Matrix

Confusion matrix adalah sebuah matrik yang digunakan untuk mengevaluasi hasil dari suatu prediksi. Menurut Han dan Kamber (2006), *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

	Prediction class 1	Prediction class 2
Actual class 1	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Actual class 2	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Nilai yang ada pada Tabel 1 menjelaskan indentifikasi dari suatu prediksi dimana *TP* dan *TN* merupakan hasil klasifikasi yang benar dari masing-masing kelas, sedangkan *FP* merupakan hasil indentifikasi yang salah dimana seharusnya data masuk pada *class 1* tapi diidentifikasi masuk ke dalam *class 2*, dan pada *FN* data yang seharusnya masuk pada *class 2* diidentifikasi masuk ke dalam *class 1*.

Adapun perhitungan akurasi pada *confusion matrix* yaitu,

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Menurut Landis dan Koch (1977) untuk menentukan suatu model klasifikasi baik, buruk, atau sedang dapat dilakukan dengan menggunakan skala nilai kappa seperti pada Tabel 2.

Tabel 2. Skala nilai kappa

Skala	Kekuatan
< 0	Buruk
0.01 – 0.2	Kurang Bagus
0.21 – 0.4	Cukup Bagus
0.41 – 0.6	Bagus
0.61 – 0.8	Sangat Bagus
0.81 - 1	Mendekati Sempurna

METODE PENELITIAN

Data dan Variabel

Data yang digunakan pada penelitian ini adalah data sekunder, yaitu data pendaftar bidikmisi UNS tahun 2018 berjumlah 2039 pendaftar. Variabel independen yang digunakan dalam penelitian ini meliputi pekerjaan ayah, pekerjaan ibu, total penghasilan orang tua, jumlah tanggungan, luas tanah, sumber air dan kepemilikan rumah. Sedangkan untuk variabel dependen yang digunakan adalah status penerimaan bidikmisi.

Metode Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah algoritma *k-NN*. Adapun langkah-langkah analisis data adalah sebagai berikut:

- 1) Melakukan normalisasi data
- 2) Membagi data menjadi data *training* dan data *testing*
- 3) Menentukan nilai *k* yang optimal
- 4) Menghitung kuadrat jarak *euclidean* objek terhadap data *training* yang diberikan
- 5) Mengurutkan hasil kuadrat jarak *euclidean* secara *ascending* (berurutan dari nilai tinggi ke rendah)
- 6) Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat diprediksikan kategori objek

HASIL DAN PEMBAHASAN

Deskripsi Data

Terdapat 7 variabel independen dan 1 variabel dependen dalam data yang digunakan dalam penelitian ini. Nilai variabel independen ditentukan sesuai dengan data mentah. Variabel independen beserta nilai variabel dapat dilihat pada Tabel 3.

Tabel 3. Penentuan Variabel dan Nilai Variabel

No.	Variabel	Nilai Variabel
1.	Penghasilan orang tua	1. > Rp 4.000.000 2. > Rp 3.500.000 – Rp 4.000.000 3. > Rp 3.000.000 – Rp 3.500.000 4. > Rp 2.500.000 – Rp 3.000.000 5. > Rp 2.000.000 – Rp 2.500.000 6. > Rp 1.500.000 – Rp 2.000.000 7. Rp 1.000.000 – Rp 1.500.000 8. < Rp 1.000.000
2.	Jumlah Tanggungan	1. 1 – 2 orang 2. 3 orang 90

		3. 4 orang
		4. 5 orang
		5. > 5 orang
3.	Pekerjaan orang tua	1. PNS 2. Pegawai Swasta 3. Wiraswasta 4. TNI / POLRI 5. Petani / Nelayan 6. Lainnya (Buruh) 7. Tidak bekerja
4.	Luas Tanah	1. > 200 m ² 2. > 100 m ² – 200 m ² 3. > 50 m ² – 100 m ² 4. > 25 m ² – 50 m ² 5. < 25 m ²
5.	Sumber Air	1. Kemasan 2. PDAM 3. Sumur 4. Sungai / mata air / gunung 5. Hujan
6.	Kepemilikan Rumah	1. Pribadi 2. Sewa Tahunan 3. Sewa Bulanan 4. Menumpang / warisan 5. Tidak memiliki

Hasil Pengolahan Data

Dalam penelitian ini digunakan 1539 data *training* dan 500 data *testing* dari calon penerima bidikmisi. *Software* RStudio digunakan untuk mengolah data dengan algoritma *k-NN*. Tabel 4 di bawah menampilkan data pengamatan mengenai calon penerima bidikmisi.

Tabel 4. Data Pengamatan

Penghasilan Orang Tua	Jumlah Tanggungan	Pekerjaan Ayah	Pekerjaan Ibu	Luas Tanah	Sumber Air	Kepemilikan Rumah	Status Penerimaan Bidikmisi
8	1	3	3	2	3	1	Ya
7	3	6	7	2	3	4	Ya
8	1	6	7	1	3	4	Ya
8	2	5	7	1	4	4	Ya
7	1	6	6	1	2	1	Ya
6	1	6	7	1	3	1	Ya
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	2	6	7	3	3	2	Ya

Tabel 5 menampilkan hasil normalisasi data. Normalisasi data ini dilakukan untuk memberikan bobot yang sama pada variabel independen yang digunakan dalam perhitungan. Langkah berikutnya setelah dilakukan normalisasi data adalah membagi data menjadi data *training* dan data *testing*, baru kemudian mencari nilai *k* yang optimal.

Tabel 5. Normalisasi Data

Penghasilan Orang Tua	Jumlah Tanggungan	Pekerjaan Ayah	Pekerjaan Ibu	Luas Tanah	Sumber Air	Kepemilikan Rumah
1.000	0.00	0.333	0.333	0.25	0.667	0.00
0.857	0.50	0.833	1.000	0.25	0.667	0.75
1.000	0.00	0.833	1.000	0.00	0.667	0.75
1.000	0.25	0.667	1.000	0.00	1.000	0.75

0.857	0.00	0.833	0.833	0.00	0.333	0.00
0.714	0.00	0.833	1.000	0.00	0.667	0.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.000	0.25	0.833	1.000	0.50	0.667	0.25

Tabel 6 menunjukkan bahwa nilai k yang memiliki tingkat akurasi dan nilai kappa terendah dan tertinggi masing-masing adalah $k=1$ dengan tingkat akurasi 80% dan nilai kappa sebesar 0.52 dan $k=5$ dengan tingkat akurasi sebesar 84.4% dan nilai kappa 0.63 yang berarti hasil klasifikasi sudah sangat bagus.

Tabel 6. Tingkat Akurasi dan Nilai Kappa Algoritma k -NN

k	Akurasi	Kappa
$k=1$	80%	0.52
$k=3$	82.4%	0.59
$k=5$	84.4%	0.63
$k=7$	83%	0.6
$k=9$	81%	0.55
$k=11$	82.8%	0.59
$k=13$	82.8%	0.59

Tabel 7 menunjukkan bahwa dari 500 data *testing*, 110 diklasifikasikan tidak mendapat bidikmisi, 49 diklasifikasikan mendapat bidikmisi tetapi ternyata tidak, 29 diklasifikasikan tidak mendapat tetapi ternyata mendapat bidikmisi, dan 312 diklasifikasikan mendapat bidikmisi. Tingkat akurasi yang diperoleh dengan menggunakan Persamaan 2 yaitu sebesar 84.4%.

Tabel 7. Confusion Matrix

Pengamatan Bidikmisi	Prediksi Bidikmisi	
	Tidak	Ya
Tidak	110	29
Ya	49	312

SIMPULAN DAN SARAN

Berdasarkan pengujian dengan menggunakan tabel perhitungan *confusion matrix* yaitu membandingkan hasil prediksi klasifikasi dengan data pengamatan bidikmisi dengan menggunakan algoritma *KNN*, dimana data yang digunakan adalah 1539 data *training* dan 500 data *testing* dapat disimpulkan bahwa hasil perhitungan yang diterapkan menghasilkan nilai akurasi sebesar 84.4% dengan nilai $k = 5$. Beberapa saran yang dapat diberikan yaitu untuk penelitian selanjutnya dapat menggunakan metode yang berbeda, seperti *decision tree*, *SVM*, *Naive Bayes Classifier*, dan metode klasifikasi lainnya untuk dapat mengetahui tingkat akurasi terbaik dari beberapa metode klasifikasi tersebut.

DAFTAR PUSTAKA

- Aulia, S., Hadiyoso, S., dan Ramadan D N. 2015. Analisis Perbandingan KNN dengan SVM untuk Klasifikasi Penyakit Diabetes Retinopati berdasarkan Citra Eksudat dan Mikroaneurisma. *Jurnal ELKOMIKA* Vol. 3 No. 1 2338-8323
- Han, J and Kamber, M. 2006. *Data Mining Concepts and Techniques, second edition*. Morgan Kaufmann, California.
- Juliardi, M S. 2017. Universitas Sebelas Maret Bidikmisi Applicant's Classification using C4.5 Algorithm. *Jurnal Ilmiah Teknologi dan Informasi* Vol. 6 No. 1 2301-7201.
- Landis, J R., & Koch, G G. 1977. *The Measurement of Observer Agreement For Categorical Data*. *Biometrics*. <https://www.jstor.org/stable/2529310>. diakses pada 18 Februari 2019.

- Liu, B. 2007. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Springer: Heidelberg, Dordrecht, London, New York.
- Mullin, M dan Suthankar, R. 2000. *Complete Cross-Validation for Nearest Neighbor Classifiers*. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.70.9106>. diunduh pada tanggal 26 Februari 2019.
- Putri R E. 2014. Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor Pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2012. *Jurnal Gaussian* Vol. 3 No. 4 Hal. 831-838.
- Ristekdikti. 2018. *Pedoman Bidikmisi 2018*. Ristekdikti, Jakarta.
- Rozaq, A. 2016. *Case Base Reasoning untuk Menentukan Kebutuhan Bahan Bangunan Rumah*. Universitas Gajah Mada, Yogyakarta.
- Tribun Jateng. 2018. <http://jateng.tribunnews.com/2018/04/17/selamat-1902-calon-mahasiswa-diterima-uns-solo-melalui-jalur-snmptn-2018>. diakses pada 18 Februari 2019.
- Yunita, D. 2017. Perbandingan Algoritma K-Nearest Neighbor dan Decision Tree untuk Penentuan Risiko Kredit Kepemilikan Mobil. *Jurnal Informatika Universitas Palembang* Vol. 2 No. 2.